

Zur Reform der Forschungsbewertung: Initiativen und Perspektiven

Matthias Koenig

1 Zur Quantifizierung in der Forschungsbewertung

Spätestens seit der Entwicklung des *Journal Impact Factors* durch das Institute for Scientific Information (ISI) vor mehr als einem halben Jahrhundert haben quantitative Indikatoren der Forschungsbewertung Einzug in das Wissenschaftssystem gehalten. Zitationsindizes finden regelmäßig Verwendung bei der Bewertung individueller Wissenschaftlerinnen oder Wissenschaftler, sind Bestandteil der ubiquitären Rankings von Universitäten und Forschungsinstituten und dienen als Hinweis auf die Förderwürdigkeit von Forschungsprojekten. Durch die seit 2000 beobachtbare rapide Digitalisierung bibliometrischer Infrastrukturen wurde die Produktion quantitativer Daten der Forschungsbewertung nochmals intensiviert (Krüger, 2000). Diese Zunahme metrischer Bewertungspraktiken ist beileibe kein Spezifikum des Wissenschaftssystems. Im Gegenteil, sie fügt sich ein in allgemeine Trends der Quantifizierung des sozialen Lebens, die soziologisch mit Begriffen des „metrischen Wir“ bzw. der „ordinal society“ erfasst wurden (Fourcade, 2016; Mau, 2017). Allerdings erzeugt sie im Wissenschaftssystem durchaus spezifische Folgeprobleme.

Diese Folgeprobleme werden seit etwa fünfzehn Jahren kritisch diskutiert und haben Anlass für eine Reihe von Initiativen gegeben, die oftmals allzu quantitativ ausgerichteten Praktiken der Forschungsbewertung zu reformieren. Zu nennen sind hier insbesondere die *San Francisco Declaration on Research Assessment* (DORA, 2012), das *Leiden Manifesto for Research Metrics* (2015) und die *Hong Kong Principles for Assessing Researchers* (2019). Die Reforminitiativen kritisieren insbesondere zwei Facetten gegenwärtiger Forschungsbewertung, die beide gleichermaßen eine Anreizstruktur erzeugen, die einer optimalen Operationsweise wissenschaftlicher Forschung abträglich ist.

Der *erste* Kritikpunkt betrifft die Dominanz quantitativer Metriken als solcher. Anstelle die substanziellen Ideen in Manuskripten, Bewerbungen oder Forschungsanträgen zu bewerten – etwa hinsichtlich origineller Fragestellungen, kreativer Theoriebildung oder innovativer Methoden – oder anstelle abzuwägen, welche konkreten Erkenntnisgewinne mit einem Forschungsprojekt erzielt wurden oder in Aussicht gestellt werden, geht es zu häufig um formale *Proxies* wie die Anzahl früherer Publikationen, um die Häufigkeit ihrer Zitationen, um den *Impact Factor* ihrer Publikationsorte oder um den Hirsch-Index ihrer Autorin bzw. ihres Autors. Diese großteils von kommerziellen

Anbietern generierten Indikatoren mögen aufgrund der Kommensurabilität numerischer Werte zwar die Bewertung von Forschung vereinfachen und beschleunigen. Sie erzeugen jedoch, wie vielfach moniert wurde, nur eine Scheinobjektivität (vgl. etwa Münch, 2015). Dass Zitationsraten und *Impact*-Faktoren je nach (sub-)disziplinärer Community variieren, dass sie mit der Anzahl von Autorinnen und Autoren korrelieren und dass bereits die Zitation eines Artikels dessen weitere Zitation – unabhängig von seiner Qualität – wahrscheinlicher macht (Matthäus-Effekt), all dies unterstreicht, wie hochgradig selektiv eine auf bibliometrische Indikatoren gestützte Bewertungspraxis tatsächlich ist. Dies ist umso problematischer, als jene Indikatoren, aber auch andere Metriken, Ratings oder Rankings sich, soweit sie in Berufungs-, Begutachtung- oder Evaluationsverfahren verankert sind oder öffentlich kommuniziert werden, unmittelbar auf die Anreizstruktur auswirken, an der sich die Akteure im Wissenschaftssystem orientieren. Diese Anreizstruktur erzeugt eine Standardisierung und Stratifikation von Forschung und führt letztlich zu einer Verarmung des wissenschaftlichen Diskurses. Man kann sogar mutmaßen, dass dies einer von mehreren Faktoren für den vielfach monierten Rückgang disruptiver Innovationen ist, durch die Forschungsprogramme und Erklärungsmodelle nicht einfach inkrementell erweitert oder ausgebessert, sondern von grundlegend neuen Paradigmen abgelöst werden (Park et al., 2023).

Eng mit dieser Kritik an bibliometrischen Indikatoren verbunden ist ein *zweiter* Punkt. In der Forschungsbewertung wird nämlich vielfach unterstellt, der einzige berücksichtigungswürdige Forschungsoutput sei die Publikation, in manchen Fachkulturen sogar nur der Fachaufsatz in möglichst hochgerankten Journalen. Mit diesem verengten Blick entgeht der Forschungsbewertung nicht nur eine große Vielfalt an Publikationsorten neben den etablierten Journalen oder (in den Geistes- und Sozialwissenschaften) den Monographie-Reihen großer englischer und US-amerikanischer Universitätsverlage. Sie übersieht auch viele alternative Arten von wissenschaftlichem Output, darunter die Erhebung von Primärdaten, die Programmierung von Forschungssoftware, die Durchführung von Replikationsstudien oder die Erschließung und Edition von Quellen. Ebenso geraten wissenschaftliche Serviceleistungen aus dem Blick, insbesondere die von Journalen, Verwaltungen und Förderorganisationen regelmäßig in Anspruch genommene Begutachtungstätigkeit, aber auch die Arbeit an digitalen Dateninfrastrukturen, die Ausbildung neuer Generationen von Forscherinnen und Forschern oder auch das wissenschaftspolitische Engagement. Die hier geleistete, oft aber unsichtbare Arbeit liefert ebenso wichtige Beiträge zur Forschung wie die in Journalen sichtbar publizierten Forschungsergebnisse. Sie in evaluativen Praktiken nicht angemessen zu berücksichtigen, erzeugt daher eine ähnlich dysfunktionale Anreizstruktur wie die Dominanz bibliometrischer Indikatoren.

2 Reforminitiativen in DFG und CoARA

Beide Kritikpunkte sind auch in der deutschen wissenschaftlichen Community bereits lange und intensiv diskutiert worden. Und tatsächlich gibt es einige Anzeichen für eine Veränderung bisher etablierter Bewertungspraktiken, so etwa in der Deutschen Forschungsgemeinschaft (DFG). Zwar kann sich die DFG nicht vorwerfen lassen, dass ihre Bewertungspraxis sich übermäßig stark auf bibliometrische *Proxies* gestützt und Publikationen priorisiert hätte. Im Gegenteil, für die Anträge auf DFG-Förderung gilt hinsichtlich von Publikationslisten bereits seit vielen Jahren das Motto „Qualität statt Quantität“, und die DFG-Entscheidungsgremien achten seit jeher darauf, dass konkrete Ideen und keine Kennziffern im Mittelpunkt der Forschungsbewertung stehen. Dennoch gibt es in der Geschäftsstelle und in den Gremien der DFG nicht nur ein großes Interesse an den laufenden Reforminitiativen, sondern auch den Willen, die eigenen Verfahren noch weiter zu optimieren. Das im Mai 2022 verabschiedete Positionspapier zum wissenschaftlichen Publizieren stellt eine scharfsichtige Analyse von Fehlentwicklungen dar und bietet sowohl der wissenschaftlichen Community als auch ihren Geldgebern konkrete Lösungsvorschläge.¹ Im September 2022 hat die DFG verbindliche Lebenslaufvorlagen eingeführt, die es erlauben – ja, die dazu ermuntern – ein breiteres Spektrum wissenschaftlicher Tätigkeiten anzugeben. Auch die Vorgaben für Publikationslisten wurden geändert, um sicherzustellen, dass verschiedene Formen des Forschungsausgangs angegeben werden können. Mit neuen Leitfäden für die Beantragung und Begutachtung von Forschungsprojekten will die DFG zudem dafür sorgen, dass die in Anträgen genannten Vorarbeiten in inhaltlichem Bezug zum Projektvorschlag stehen, anstatt bloß aufgezählt zu werden.²

Dies ist allerdings nicht alles. Die DFG hat sich außerdem aktiv in den von der Europäischen Kommission angestoßenen Diskussionen zur Reform der Forschungsbewertung engagiert. Letztere haben 2022 zur Verabschiedung des *Agreement on Reforming Research Assessment* geführt, zu dessen Erstunterzeichnerinnen die DFG gehörte.³ Im Hintergrund dieses Engagements stand die Überlegung, dass eine Veränderung von Praktiken der Forschungsbewertung weit mehr erfordert als die Bemühungen einzelner Organisationen. Denn die oben geschilderten Fehlanreize etablierter Bewertungspraktiken sind in ein komplexes System gegenseitiger Erwartungen vieler Individuen, Organisationen und Institutionen eingebettet und damit Teil einer *Bewertungskultur* mit hohen Beharrungskräften. Die Persistenz der viel

¹DFG (2022). Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung. Herausforderungen und Handlungsfelder. Abgerufen am 28.06.2023 unter https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/publikationswesen/

²Siehe dazu DFG (2022). Information für die Wissenschaft Nr. 61. Abgerufen am 28.06.2023 unter https://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_61/

³Siehe dazu https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf (abgerufen am 28.06.2023).

kritisierten Universitätsrankings zeigt dies mit besonderer Deutlichkeit (Hamann & Ringel, 2023). Praktiken der Forschungsbewertung lassen sich dementsprechend nur im Rahmen eines allgemeinen *Kulturwandels* beeinflussen und ändern.

Um einen solchen Kulturwandel anzustoßen und nachhaltig zu gestalten, wurde auf Grundlage des *Agreements* im Dezember 2022 die *Coalition for Advancing Research Assessment* (CoARA) gegründet.⁴ Die CoARA verpflichtet sich auf einige übergeordnete Ziele bzw. *Core Commitments*, darunter insbesondere die Berücksichtigung der Diversität von Forschungsbeiträgen und die Abkehr von der unreflektierten Nutzung quantitativer Metriken und Rankings zugunsten peer-review-gestützter qualitativer Forschungsbewertung. Ihr Arbeitsmodus besteht dabei im Austausch über mögliche Reformschritte, aber auch in der kontinuierlichen Arbeit an der Präzisierung der übergeordneten Ziele. Was *genau* die „optimale“ Funktionsweise wissenschaftlicher Forschung gewährleistet, wird dabei ebenso diskutiert wie die Frage, wie *genau* Praktiken der Forschungsbewertung in spezifischen nationalen und disziplinären Kontexten und in spezifischen (kompetitiven vs. nicht-kompetitiven; einmaligen vs. wiederholten) Bewertungssituationen zu reformieren sind. Schon gemessen an ihren Mitgliederzahlen kann die Initiative im ersten Jahr nach ihrer Gründung als Erfolg bezeichnet werden. Inzwischen gehören über 500 Organisationen – darunter Universitäten, Forschungsinstitute und Forschungsfördereinrichtungen – aus 44 Ländern der CoARA an, 580 Organisationen haben gar das *Agreement* unterzeichnet. Aktuell starten die ersten Arbeitsgruppen, die sich einzelnen Themen wie beispielsweise der Bewertung von Forschungsprojekten, der Bewertung akademischer Karrieren oder auch der angemessenen Anerkennung von *Peer Review* widmen, und Synergien mit anderen Reforminitiativen der Forschungsbewertung (z. B. DORA) beginnen Gestalt anzunehmen.

3 Kritische Nachfragen und erste Antworten

Bei aller Zuversicht werden jedoch auch kritische Fragen an die CoARA gerichtet, nicht zuletzt in Deutschland:

1. Ist der Status quo der Forschungsbewertung wirklich so problematisch?
2. Birgt die CoARA nicht die Gefahr, dass bei der Forschungsbewertung das Kriterium der Exzellenz aus dem Blick gerät?
3. Und unterminiert die CoARA aufgrund ihrer Entstehungsgeschichte womöglich die Autonomie von Wissenschaftlerinnen und Wissenschaftlern, Forschungsorganisationen oder gar des deutschen Wissenschaftssystems als Ganzem?

⁴Siehe www.coara.eu (abgerufen am 28.06.2023).

Die kontroverse Diskussion unterstreicht das allgemeine Interesse am Thema Forschungsbewertung und ist insofern unbedingt zu begrüßen. Zugleich scheint mir, dass hinter den Fragen auch einige Missverständnisse stehen, die im Interesse zukünftiger Entwicklungen des Wissenschaftssystems ausgeräumt werden sollten.

Zunächst zur *ersten* Frage: Steht es wirklich überall so schlimm um die aktuelle Bewertungspraxis? Zeigt das oben genannte Beispiel der DFG nicht, dass quantitative Metriken in Deutschland weniger verbreitet sind als in anderen nationalen Wissenschaftssystemen? Und variiert der Reformdruck nicht auch nach Fachkulturen und Bewertungssituationen? In doppelter Hinsicht scheint mir diese Frage die Problemlage zu unterschätzen. Zum einen ist das moderne Wissenschaftssystem hochgradig transnational verflochten, sodass Bewertungspraktiken sich nur schwerlich in nationaler Isolation bestimmen lassen. Forscherinnen und Forscher, aber auch Hochschulen und Forschungsinstitute sind schlechterdings gezwungen, in ihren strategischen Entscheidungen – vom Forschungsprojekt bis hin zur Publikation – auch internationale Erwartungen und Bewertungsstandards zu berücksichtigen. Zum anderen wird auch in Deutschland, insbesondere seitens der jüngeren Generation, durchaus noch Reformbedarf hinsichtlich der Forschungsbewertung gesehen.⁵ Es gibt also mehrere gute Gründe, die laufenden internationalen Reforminitiativen im Rahmen von CoARA auch von deutscher Seite nicht nur zu unterstützen, sondern aktiv mitzugestalten.

Mit der *zweiten* Frage verbindet sich die Sorge, die Reform der Forschungsbewertung könne das gerade in Deutschland gern zitierte Kriterium der „Exzellenz“ auf die eine oder andere Weise aufweichen. Manche suggerieren, gerade diejenigen forderten eine Reform, die in der dünnen Luft der „wirklich“ exzellenten Wissenschaft nicht wettbewerbsfähig und damit letztlich befangen seien. Diesem Verdacht der Befangenheit lässt sich allerdings leicht entgegnen, dass auch diejenigen, die im bestehenden Bewertungssystem mit seinen verschiedenen Exzellenzsiegeln erfolgreich sind, *vested interests* haben. Wer bei metrischen Indikatoren hoch punktet, ist womöglich weniger geneigt zu konzedieren, wie problematisch die Verwendung bibliometrischer *Proxies* ist. Andere suggerieren wiederum, dass unter dem Segel von CoARA wissenschaftliche Sekundärkriterien – etwa Diversität – die Primärkriterien von Forschungsqualität überlagern könnten. Auch diese Sorge ist meines Erachtens weitgehend unbegründet. Wenn im *Agreement* von Diversität die Rede ist, so geht es in allererster Linie um die Anerkennung der oben angesprochenen Vielfalt von Beiträgen zum Forschungssystem im Rahmen eines holistischen Exzellenzverständnisses. Zwar wird im CoARA-Kontext ebenfalls häufig die Position vertreten, dass die Forschungsqualität auch von ausgewogenen Geschlechterverhältnissen und sonstiger Diversität profitieren kann. Der

⁵Vgl. z.B. Die Junge Akademie (2020). Anreiz-Problematiken in der Wissenschaft. Stellungnahme vom 22.09.2020. Abgerufen am 28.06.2023 unter <https://www.diejungeakademie.de/de/publikationen/anreiz-problematiken-in-der-wissenschaft>

Fokus auf höchste Qualitätsansprüche als Primärkriterium der Forschungsbewertung wird durch diese Position jedoch keineswegs geschwächt, sondern eher noch gestärkt. Um es pointiert zu sagen: Das *Agreement* stellt keineswegs die Prinzipien exzellenter Forschung infrage. Gewiss, der Text verzichtet auf die explizite Nennung des Terminus „Exzellenz“ (wofür es angesichts seiner inflationären Verwendung in Hochschulrankings im Übrigen gute Gründe gibt). Aber es ist völlig unstrittig, dass das *Agreement* und mit ihm auch CoARA gerade die Verbesserung der Verfahren zur Identifikation von Forschungsleistung und damit auch von Spitzenforschung anvisieren.

Die *dritte* Frage schließlich betrifft die mögliche Gefahr für die Autonomie des deutschen Wissenschaftssystems. Motiviert ist sie durch die Wahrnehmung, dass die CoARA ihren Ursprung in Amtsstuben der Europäischen Kommission nahm. Und tatsächlich hatte letztere die Federführung der ersten Reformpapiere inne,⁶ führte den Konsultationsprozess zum *Agreement* durch und begleitet CoARA bis heute als wissenschaftspolitisches Schwergewicht. Es ist aber wichtig zu betonen, dass die Kommission bei der Ausarbeitung der CoARA-Ziele in der *Drafting Group* eng mit wichtigen Dachverbänden – der European University Association (EUA) und Science Europe (SE) – und später mit vielen weiteren Organisationen in der *Core* und der *Implementation Group* zusammenarbeitete. Inzwischen operiert CoARA durch und durch bottom-up, was allen Mitgliedsorganisationen über die Generalversammlung und die themenspezifischen Arbeitsgruppen die Möglichkeit eröffnet, die Reformbestrebungen der Forschungsbewertung aktiv mitzugestalten. Vor diesem Hintergrund sollte die wissenschaftliche Community es eher begrüßen, dass sich die Kommission so interessiert an den Reformdiskussionen innerhalb des Wissenschaftssystems zeigt.

4 **Ausblick**

Schließen möchte ich mit einigen Überlegungen zu zukünftigen Herausforderungen und Chancen einer Reform der Forschungsbewertung. Zunächst zu den Herausforderungen: So begrüßenswert die Abkehr von ausschließlich quantitativ basierten Bewertungen ist, so wenig wird man auf diese – u. a. aufgrund des personellen und organisatorischen Aufwands qualitativer Bewertungen – vollständig verzichten können. Und so berechtigt das Anliegen ist, der Diversität von Forschungsausgaben Rechnung zu tragen, so sehr ist darauf zu achten, dass neue Praktiken der Forschungsbewertung keine Flut neuer ordinaler Indikatoren erzeugen, die ihrerseits wieder Fehlanreize setzen und reaktive Anpassungsstrategien wissenschaftlicher Akteure erzeugen können. Dies gilt insbesondere für den Aspekt des *research impact*, den manche Beteiligte am Reformprozess gern in den Rang eines Primärkriteriums der Forschungsbewertung

⁶Vgl. z. B. European Commission Communication (2020). A New European Research Area for Research and Innovation. EC Commission Communication 628, 30. September 2020. Abgerufen am 28.06.2023 unter <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0628&qid=1614808291158>.

erheben möchten, damit aber womöglich den intrinsischen Wert erkenntnisorientierter Grundlagenforschung unterschätzen und Einfallstore für die politische Steuerung wissenschaftlicher Themenfindung schaffen. Vor allem aber darf bei allem Reformeifer nicht übersehen werden, dass die zunehmende Frequenz, Intensität und Bürokratisierung der Forschungsbewertung (Stichwort „Evaluitis“) als solche dem in langfristigen Zyklen operierenden Wissenschaftssystem abträglich sein könnten.

Gleichwohl ist die Stoßrichtung der aktuellen Reforminitiativen, darunter auch der CoARA, weiterhin unbedingt zu bestärken. Im Idealfall verbessern diese Initiativen die Rahmenbedingungen und Anreizstrukturen für disruptive Innovationen, die angesichts vielfältiger globaler Herausforderungen so dringlich sind wie selten zuvor. In jedem Fall aber fördern sie eine Rückbesinnung auf das Proprium wissenschaftlichen Forschens, nämlich die zugleich leidenschaftliche und methodische Suche nach neuer Erkenntnis, die nach kritischer Prüfung im wissenschaftlichen Diskurs unsere gemeinsamen Wissensspeicher erweitert und mithin zu unserer Orientierungs- und Handlungsfähigkeit in der Welt beiträgt.

Literatur

- Fourcade, M. (2016). Ordinalization. *Sociological Theory*, 34(3), 175–195.
- Hamann, J. & Ringel, L. (2023). The discursive resilience of university rankings. *Higher Education*. <https://doi.org/10.1007/s10734-022-00990-x>
- Krüger, A. (2000). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, 8(2), 58–67.
- Mau, S. (2017). *Das Metrische Wir. Über die Quantifizierung des Sozialen*. Suhrkamp.
- Münch, R. (2015). Alle Macht den Zahlen! Zur Soziologie des Zitationsindexes. *Soziale Welt*, 66, 149–149.
- Park, M., Leahey, E. & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613, 138–144.

Angaben zum Autor:

Prof. Dr. Matthias Koenig
Ruprecht-Karls-Universität Heidelberg
Max-Weber-Institut für Soziologie
Bergheimer Straße 58
69115 Heidelberg

Matthias Koenig ist Professor für Empirische Makrosoziologie an der Universität Heidelberg und seit 2021 Vizepräsident der Deutschen Forschungsgemeinschaft (DFG).